Journal of Computational Mathematics
Vol.35, No.4, 2017, 529–546.
http://www.global-sci.org/jcm
doi:10.4208/jcm.1702-m2016-0699

EXTENDED LEVENBERG-MARQUARDT METHOD FOR COMPOSITE FUNCTION MINIMIZATION*

Jianchao Huang
School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China
Email: huangjc@sjtu.edu.cn
Zaiwen Wen
Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China
Email: wenzw@pku.edu.cn
Xiantao Xiao
School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China
Email: xtxiao@dlut.edu.cn

Abstract

In this paper, we propose an extended Levenberg-Marquardt (ELM) framework that generalizes the classic Levenberg-Marquardt (LM) method to solve the unconstrained minimization problem $\min \rho\,(r(x))$, where $r: \mathbb{R}^n \to \mathbb{R}^m$ and $\rho: \mathbb{R}^m \to \mathbb{R}$. We also develop a few inexact variants which generalize ELM to the cases where the inner subproblem is not solved exactly and the Jacobian is simplified, or perturbed. Global convergence and local superlinear convergence are established under certain suitable conditions. Numerical results show that our methods are promising.

Mathematics subject classification: 65K05, 90C30, 90C53.
Key words: Unconstrained minimization, Composite function, Levenberg-Marquardt method.

1. Introduction

Many applications can be written as minimization problems of the form

$$\min_x \ \phi(x) := \rho\,(r(x))\,, \tag{1.1}$$

where $r: \mathbb{R}^n \to \mathbb{R}^m$ and $\rho: \mathbb{R}^m \to \mathbb{R}$. The function r is often referred to as the residual or modeling error, and ρ the penalty function of residual r. When $\rho(u) = \frac{1}{2}\|u\|^2$, (1.1) is the well known nonlinear least squares problem

$$\min_x \frac{1}{2}\|r(x)\|^2. \tag{1.2}$$

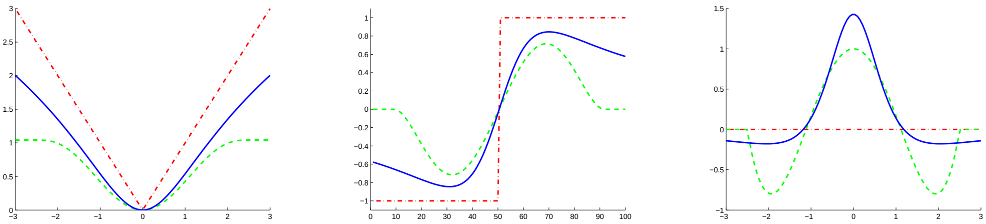
Throughout this paper, we denote $\|\cdot\|$ as the L_2-norm. Another commonly seen example is that

$$\min_x \sum_{i=1}^m \rho_i(r_i(x)), \tag{1.3}$$

where $\rho(u)$ takes the form

$$\rho(u) = \rho_1(u_1) + \cdots + \rho_m(u_m),$$

* Received July 6, 2016 / Revised version received October 19, 2016 / Accepted February 28, 2017 / Published online June 1, 2017 /



(a) $\rho(r)$ (b) $\nabla\rho(r)$ (c) $\nabla^2\rho(r)$

Fig. 1.1. Penalty function, its gradient, and Hessian of some common models: Laplace (red dash-dot), Tukey (green dash), and Student's $t-$ (blue line).

and $\rho_i : \mathbb{R} \to \mathbb{R}$ is a function of scalar variables. Some examples of ρ are illustrated in Figure 1.1. More frequently used penalty functions are summarized in Table 6.1.

A general method that solves the unconstrained minimization problem (1.1) is Newton's method. Let $f_k := f(x_k)$ be a function f evaluated at point x_k. At each iteration, the method finds a direction by solving the quadratic approximation problem

$$\min_d \ \phi_k + g_k^T d + \frac{1}{2} d^T H_k d, \tag{1.4}$$

where $g_k := \nabla\phi_k$ and $H_k := \nabla^2\phi_k$ are the gradient and Hessian of ϕ. It is known that this method has fast local convergence. Since the cost to evaluate the Hessian H_k may be expensive in general, methods using only the gradient information are preferred.

When $\rho(u) = \frac{1}{2}\|u\|^2$, the problem (1.1) degenerates to the classic nonlinear least squares problem (1.2). The Levenberg-Marquardt (LM) method [13,16] can be regarded as a regularized Gauss-Newton method that solves the nonlinear least squares problem (1.2) iteratively by a sequence of linear least squares problems

$$\min_d \ \frac{1}{2}\left\|r_k + J_k d\right\|^2 + \frac{\tau_k}{2}\left\|d\right\|^2, \tag{1.5}$$

where $J(x) := \nabla r(x)^T$ is the Jacobian matrix and $\tau_k > 0$ is some positive constant. The Hessian H_k in (1.4) is approximated by the positive semidefinite matrix $J_k^T J_k$. Since $J_k = \nabla r_k^T$ is already contained in the first-order derivative, we obtain $\nabla^2\phi_k$ almost for free.

The LM method is more computationally attractive compared to the Gauss-Newton method or the Newton's method since it avoids singularity and does not evaluate the second order derivative. It also enjoys a quadratic convergent rate when $\tau_k = \|r_k\|^p$ for $p \in [1,2]$, see, e.g., [10, 24]. These advantages motivate us to extend the LM method to the general problem (1.1).

For large scale applications, it is impractical to solve the problem (1.4) exactly. An iterative approach such as the conjugate gradient (CG) method is more favorable. However, the CG method is designed to solve positive definite systems, and H_k may have negative eigenvalues when the iterate is not close to the solution. The truncated CG method ([22], for example) terminates the CG iteration as soon as negativity is detected. Usually, it is also equipped with a line search or trust region method for better global convergence.

In this paper, we generalize the classic LM method to solve the composite function minimization problem (1.1). A weighted least squares method is incorporated to simplify the approximated Hessian and accelerate the computation at initial steps when the residuals are